Empowering Users Through Technical Transparency

– Simon Althaus

PI-Tandem: Prof. Dr. Max Mühlhäuser / Prof. Dr. Peter Buxmann
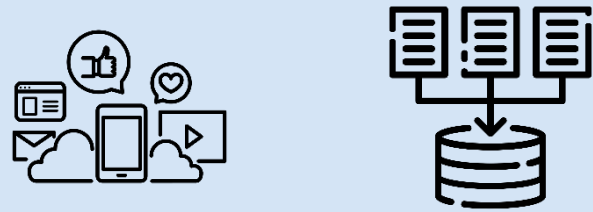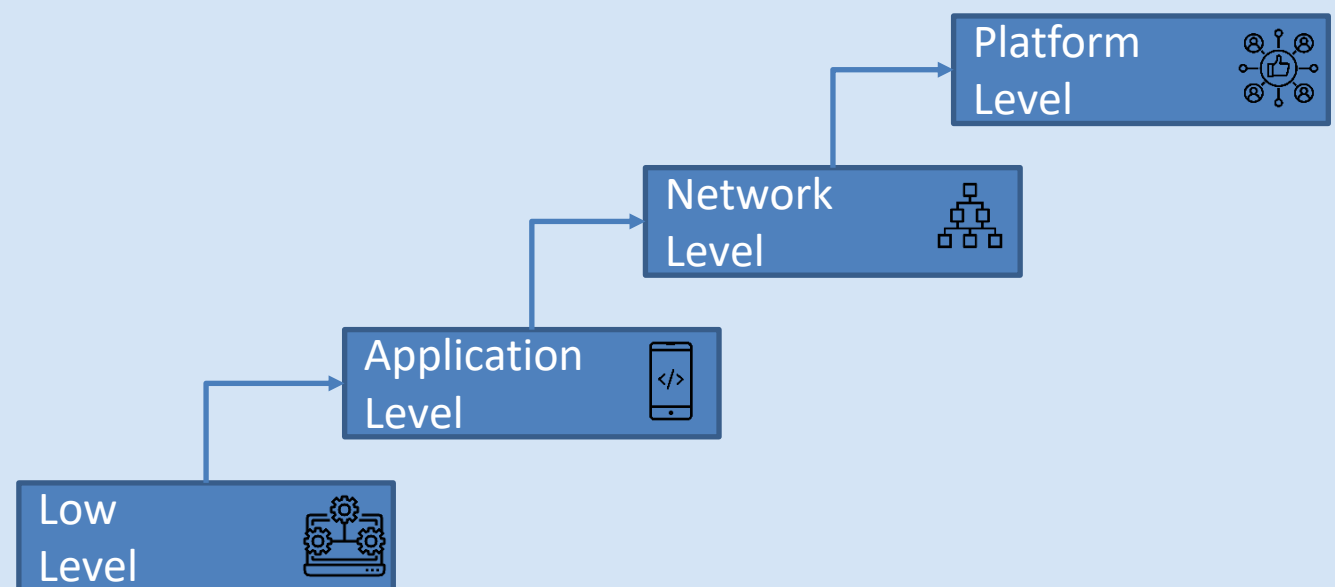
B.2.

Tandem with A.1 and A.3

Motivation

- Platform providers of large Internet services are known to collect large amounts of data on their users
  - Used to monetize and provide targeted advertising for example
- Data collection of such platforms is often
  - Neither in the best interest of the users
  - Nor are users aware of the extend and impact
- Smartphones and mobile apps of respective platforms widely used and have access to sensitive information

Related Work

- Possibility to request data a provider has on the user based on the right of access (GDPR)
  - Provides first insights, but not applicable outside EU and not guaranteed to be complete
- **Transparency Enhancing Tools (TETs)** – try to make the user aware of this data collection, e.g., on mobile devices
- Previous work mainly focused on
  - The network view of data collection instead of directly looking at the device level
  - Discrepancies between network data collection and privacy policy statements
  - Potential data collection as identified by permissions of applications
- ➔ Current approaches like TETs are still lacking in
  - The completeness of information gathered
  - Illustrating implications of what platform providers can do with users' data
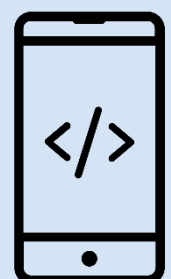
Research Overview

- **Goal:** Empower users by increasing the transparency of this data collection by platform providers on mobile devices
- ➔ Build a multi-faceted Transparency Enhancing Tool (TET)
- **Expected benefits**
  - More complete view of the gathered information on users
  - Potential to provide explanations on observations how derived data is used, e.g., for targeted advertising

Research Approach

- Build a multi-faceted Transparency Enhancing Tool (TET), utilizing information from different levels:

1. **Low Level**
   - Discover what data is collection on a low level of the operating system
   - Whether this data collection is (ab)normal
   - With whom this information is potentially shared
     ➔ information flow analysis
   - Starting point: adopting an approach from the insider detection domain
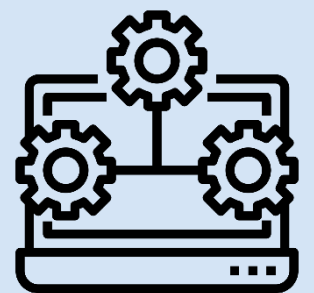
2. **Application Level**
   - Detect explicit and implicit data gathering of Android applications
   - Create shadow profiles that depict what information a service provider has collected from a user
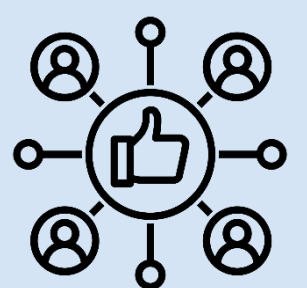   - Potentially infer additional information like preferences

3. **Network Level**
   - Data may be used to complement information discovered by previous levels or to compare completeness
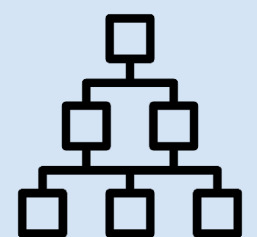
4. **Platform Level**
   - Monitor users' recommendations
   - Combine information (shadow profiles) of different users
   - Detection correlations or anomalies between multiple users
   - Derive what additional information can be induced that was not apparent beforehand
   - ➔ Simulate the platform provider's view to reason about actual platform behavior